# On the monotonicity of $Q^2$ spectral element method for Laplacian on quasi-uniform rectangular meshes

Logan J. Cross[1] and Xiangxiong Zhang[1,*],

[1] *Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067.*

**Abstract.** The monotonicity of discrete Laplacian implies discrete maximum principle, which in general does not hold for high order schemes. The $Q^2$ spectral element method has been proven monotone on a uniform rectangular mesh. In this paper we prove the monotonicity of the $Q^2$ spectral element method on quasi-uniform rectangular meshes under certain mesh constraints. In particular, we propose a relaxed Lorenz's condition for proving monotonicity.

**AMS subject classifications**: 65N30, 65N06, 65N12

**Key words**: Inverse positivity, discrete maximum principle, high order accuracy, monotonicity, discrete Laplacian, quasi uniform meshes, spectral element method

## 1 Introduction

In many applications, monotone discrete Laplacian operators are desired and useful for ensuring stability such as discrete maximum principle or positivity-preserving of physically positive quantities [6, 10, 18, 21]. Let $\Delta_h$ denote the matrix representation of a discrete Laplacian operator, then it is called *monotone* if $(-\Delta_h)^{-1} \geq 0$, i.e., the inverse matrix $(-\Delta_h)^{-1}$ has nonnegative entries. In this paper, all inequalities for matrices are entry-wise inequalities.

In the literature, the most important tool for proving monotonicity is via nonsingular M-matrices, which are inverse-positive matrices. See the Appendix for a convenient characterization of the M-matrices. The simplest second order accurate centered finite difference $u''(x_i) \approx \frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1})}{\Delta x^2}$ is monotone because the corresponding matrix $(-\Delta_h)^{-1}$ is an M-matrix thus inverse positive. Even though the linear finite element method forms an M-matrix on unstructured triangular meshes under a mild mesh constraint [24], in general the discrete maximum principle is not true for high order finite element methods on unstructured meshes [9]. On the other hand, there exist a few high order accurate inverse positive schemes on structured meshes.

---

*Corresponding author. *Email addresses:* `logancross68@gmail.com` (Cross), `zhan1966@purdue.edu` (Zhang)

For solving a Poisson equation, provably monotone high order accurate schemes on structured meshes include the classical 9-point scheme [3, 7, 11] in which the stiffness matrix is an M-matrix. The classical 9-point scheme has the same stiffness matrix as fourth order accurate compact finite difference schemes [13], see the appendix in [16]. In [2, 4], a fourth order accurate finite difference scheme was constructed and its stiffness matrix is a product of two M-matrices thus monotone. The Lagrangian $P^2$ finite element method on a regular triangular mesh [23] has a monotone stiffness matrix [19]. On an equilateral triangular mesh, the discrete maximum principle of $P^2$ element can also be proven [9]. Monotonicity was also proven for the $Q^2$ spectral element method on an uniform rectangular mesh for a variable coefficient Poisson equation under suitable mesh constraints [14]. The $Q^k$ spectral element method is the continuous finite element method with Lagrangian $Q^k$ basis implemented by $(k+1)$-point Gauss-Lobatto quadrature. The monotonicity of $Q^3$ spectral element method for Laplacian on uniform meshes was also proven in [8].

For proving inverse positivity, the main viable tool in the literature is to use M-matrices which are inverse positive. A convenient sufficient condition for verifying the M-matrix structure is to require that off-diagonal entries must be non-positive. Except the fourth order compact finite difference, all high order accurate schemes induce positive off-diagonal entries, destroying M-matrix structure, which is a major challenge of proving monotonicity. In [2] and [1], and also the appendix in [14], M-matrix factorizations of the form $(-\Delta_h)^{-1} = M_1 M_2$ were shown for special high order schemes but these M-matrix factorizations seem ad hoc and do not apply to other schemes or other equations. In [19], Lorenz proposed some matrix entry-wise inequality for ensuring a matrix to be a product of two M-matrices and applied it to $P^2$ finite element method on uniform regular triangular meshes.

In [14], Lorenz's condition was applied to $Q^2$ spectral element method on uniform rectangular meshes. Such a monotonicity result implies that the $Q^2$ spectral element method is bound-preserving or positivity-preserving for convection diffusion equations including the Allen-Cahn equation [21], the Keller-Segel equation [10], the Fokker-Planck equation [17], as well as the internal energy equation in compressible Navier-Stokes system [18]. On the other hand, all these results about $Q^2$ spectral element method are on uniform meshes. For both theoretical and practical interests, a natural question to ask is whether such a monotonicity result still holds on non-uniform meshes. The monotonicity of high order schemes on quasi-uniform meshes are preferred in many applications, e.g., [22].

The focus of this paper is to discuss Lorenz's condition for $Q^2$ spectral element method on quasi-uniform meshes. We discuss and derive sufficient mesh constraints to preserve monotonicity of $Q^2$ spectral element method on a quasi-uniform rectangular mesh. In general, the same discussion also applies to Lagrangian $P^2$ finite element method on a quasi-uniform regular triangular mesh, but there does not seem to be any advantage of using $P^2$.

For simplicity, we will focus only on Dirichlet boundary conditions. For Neumann

boundary conditions, the discussion of monotonicity is very similar, e.g., see [10, 17] for discussion on Neumann boundaries.

The rest of the paper is organized as follows. In Section 2, we briefly review the $Q^2$ spectral element method and its equivalent finite difference form for the Poisson equation. In Section 3, we review the Lorenz's condition for proving monotonicity and propose a relaxed version of Lorenz's condition. Though we only focus on $Q^2$ spectral element method on quasi-uniform meshes for Laplacian in this paper, the proposed relaxed Lorenz's condition may also be used to derive monotonicity under more relaxed mesh constraints for $Q^2$ spectral element method solving variable coefficient problems such as those in [10, 14, 17]. In Section 4, we prove the monotonicity of $Q^2$ spectral element method on a quasi-uniform mesh by using the relaxed Lorenz's condition. Numerical tests of accuracy of the scheme and necessity of the mesh constraints for monotonicity are given in Section 5. Section 6 are concluding remarks.

# 2 $Q^2$ spectral element method

## 2.1 Finite element method with the simplest quadrature

Consider an elliptic equation on $\Omega = (0,1) \times (0,1)$ with Dirichlet boundary conditions:

$$\mathcal{L}u \equiv -\nabla \cdot (a\nabla u) + cu = f \quad \text{on} \quad \Omega, \quad u = g \quad \text{on} \quad \partial\Omega. \tag{2.1}$$

Assume there is a function $\bar{g} \in H^1(\Omega)$ as an extension of $g$ so that $\bar{g}|_{\partial\Omega} = g$. The variational form of (2.1) is to find $\tilde{u} = u - \bar{g} \in H_0^1(\Omega)$ satisfying

$$\mathcal{A}(\tilde{u}, v) = (f, v) - \mathcal{A}(\bar{g}, v), \quad \forall v \in H_0^1(\Omega), \tag{2.2}$$

where $\mathcal{A}(u, v) = \iint_\Omega a\nabla u \cdot \nabla v \, dx dy + \iint_\Omega cuv \, dx dy$, $(f, v) = \iint_\Omega f v \, dx dy$.
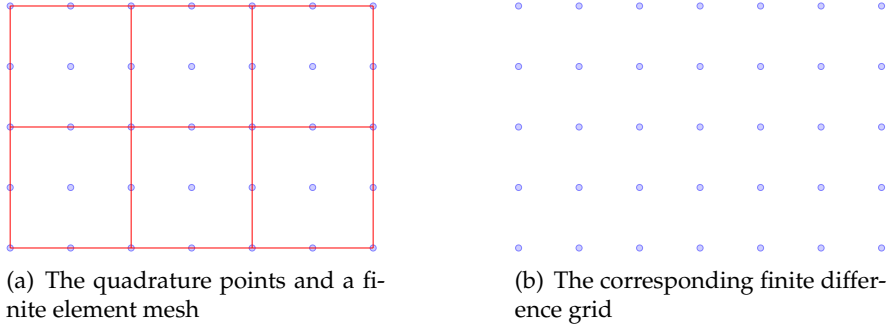


(a) The quadrature points and a finite element mesh

(b) The corresponding finite difference grid

Figure 1: An illustration of Lagrangian $Q^2$ element and the $3 \times 3$ Gauss-Lobatto quadrature.

Let $h$ be quadrature point spacing of a rectangular mesh shown in Figure 1 and $V_0^h \subseteq H_0^1(\Omega)$ be the continuous finite element space consisting of $Q^2$ polynomials, then the

most convenient implementation of finite element method is to use the simple quadrature consisting of $3 \times 3$ Gauss-Lobatto quadrature rule for all the integrals, see Figure 1 for $Q^2$ method. Such a numerical scheme can be defined as: find $u_h \in V_0^h$ satisfying

$$\mathcal{A}_h(u_h, v_h) = \langle f, v_h \rangle_h - \mathcal{A}_h(g_I, v_h), \quad \forall v_h \in V_0^h, \tag{2.3}$$

where $\mathcal{A}_h(u_h, v_h)$ and $\langle f, v_h \rangle_h$ denote using simple quadrature for integrals $\mathcal{A}(u_h, v_h)$ and $(f, v_h)$ respectively, and $g_I$ is the piecewise $Q^2$ Lagrangian interpolation polynomial at the quadrature points shown Figure 1 of the following function:

$$g(x,y) = \begin{cases} 0, & \text{if } (x,y) \in (0,1) \times (0,1), \\ g(x,y), & \text{if } (x,y) \in \partial\Omega. \end{cases}$$

Then $\bar{u}_h = u_h + g_I$ is the numerical solution for the problem (2.1). Notice that (2.3) is not a straightforward approximation to (2.2) since $\bar{g}$ is never used. When the numerical solution is represented by a linear combination of Lagrangian interpolation polynomials at the grid points, it can be rewritten as a finite difference scheme. We can also call it a variational difference scheme since it is derived from the variational form.

## 2.2 The difference formulation

The scheme (2.3) with Lagrangian $Q^2$ basis can also be written as a finite difference scheme [15].

Consider a uniform grid $(x_i, y_j)$ for a rectangular domain $[0,1] \times [0,1]$ where $x_i = ih$, $i = 0,1,\ldots,n+1$ and $y_j = jh$, $j = 0,1,\ldots,n+1$, $h = \frac{1}{n+1}$, where $n$ must be odd. Let $u_{ij}$ denote the numerical solution at $(x_i, y_j)$. Let $\mathbf{u}$ denote an abstract vector consisting of $u_{ij}$ for $i, j = 1, 2, \cdots, n$. Let $\bar{\mathbf{u}}$ denote an abstract vector consisting of $u_{ij}$ for $i, j = 0, 1, 2, \cdots, n, n+1$. Let $\bar{\mathbf{f}}$ denote an abstract vector consisting of $f_{ij}$ for $i, j = 1, 2, \cdots, n$ and the boundary condition $g$ at the boundary grid points. Then the matrix vector representation of (2.3) is $S\bar{\mathbf{u}} = M\mathbf{f}$ where $S$ is the stiffness matrix and $M$ is the lumped mass matrix. For convenience, after inverting the mass matrix, with the boundary conditions, the whole scheme can be represented in a matrix vector form $\bar{L}_h \bar{\mathbf{u}} = \bar{\mathbf{f}}$. For Laplacian $\mathcal{L}u = -\Delta u$, $\bar{L}_h \bar{\mathbf{u}} = \bar{\mathbf{f}}$ on a uniform

mesh is given as

$$(\bar{L}_h \bar{\mathbf{u}})_{i,j} := \frac{-u_{i-1,j} - u_{i+1,j} + 4u_{i,j} - u_{i,j+1} - u_{i+1,j}}{h^2} = f_{i,j}, \quad \text{if } (x_i, y_j) \text{ is a cell center,}$$

$$(\bar{L}_h \bar{\mathbf{u}})_{i,j} := \frac{-u_{i-1,j} + 2u_{i,j} - u_{i+1,j}}{h^2} + \frac{u_{i,j-2} - 8u_{i,j-1} + 14u_{i,j} - 8u_{i,j+1} + u_{i,j+2}}{4h^2} = f_{i,j},$$
$$\text{if } (x_i, y_j) \text{ is an edge center for an edge parallel to the x-axis,}$$

$$(\bar{L}_h \bar{\mathbf{u}})_{i,j} := \frac{u_{i-2,j} - 8u_{i-1,j} + 14u_{i,j} - 8u_{i+1,j} + u_{i+2,j}}{4h^2} + \frac{-u_{i,j-1} + 2u_{i,j} - u_{i,j+1}}{h^2} = f_{i,j},$$
$$\text{if } (x_i, y_j) \text{ is an edge center for an edge parallel to the y-axis,}$$

$$(\bar{L}_h \bar{\mathbf{u}})_{i,j} := \frac{u_{i-2,j} - 8u_{i-1,j} + 14u_{i,j} - 8u_{i+1,j} + u_{i+2,j}}{4h^2} + \frac{u_{i,j-2} - 8u_{i,j-1} + 14u_{i,j} - 8u_{i,j+1} + u_{i,j+2}}{4h^2} = f_{i,j},$$
$$\text{if } (x_i, y_j) \text{ is a knot,}$$

$$(\bar{L}_h \bar{\mathbf{u}})_{i,j} := u_{i,j} = g_{i,j} \quad \text{if } (x_i, y_j) \text{ is a boundary point.}$$

$$(2.4)$$

If ignoring the denominator $h^2$, then the stencil can be represented as:

$$
\text{cell center} \quad
\begin{matrix}
 & -1 & \\
-1 & 4 & -1 \\
 & -1 &
\end{matrix}
\qquad
\text{knots} \quad
\begin{matrix}
 & & \frac{1}{4} & & \\
 & & -2 & & \\
\frac{1}{4} & -2 & 7 & -2 & \frac{1}{4} \\
 & & -2 & & \\
 & & \frac{1}{4} & &
\end{matrix}
$$

$$
\text{edge center (edge parallel to } y\text{-axis)} \quad
\begin{matrix}
 & & -1 & & \\
\frac{1}{4} & -2 & \frac{11}{2} & -2 & \frac{1}{4} \\
 & & -1 & &
\end{matrix}
$$

$$
\text{edge center (edge parallel to } x\text{-axis)} \quad
\begin{matrix}
 & \frac{1}{4} & \\
 & -2 & \\
-1 & \frac{11}{2} & -1 \\
 & -2 & \\
 & \frac{1}{4} &
\end{matrix}
\quad .
$$

**Remark 2.1.** When regarded as a finite difference scheme, the scheme (2.3) is fourth order accurate in $\ell^2$-norm for elliptic, parabolic, wave and Schrödinger equations [12, 15].

# 3 Lorenz's condition for monotonicity

In this section, we first review the Lorenz's method for proving monotonicity [19], then present a relaxed Lorenz's condition. The definition of M-matrices is given in the appendix.

## 3.1 Discrete maximum principle

We first review how the monotonicity implies the discrete maximum principle for a boundary value problem. For a finite difference scheme, assume there are $N$ grid points in the domain $\Omega$ and $N^\partial$ boundary grid points on $\partial\Omega$. Define

$$\mathbf{u} = \begin{pmatrix} u_1 & \cdots & u_N \end{pmatrix}^T, \mathbf{u}^\partial = \begin{pmatrix} u_1^\partial & \cdots & u_{N^\partial}^\partial \end{pmatrix}^T, \tilde{\mathbf{u}} = \begin{pmatrix} u_1 & \cdots & u_N & u_1^\partial & \cdots & u_{N^\partial}^\partial \end{pmatrix}^T.$$

A finite difference scheme can be written as

$$\mathcal{L}_h(\tilde{\mathbf{u}})_i = \sum_{j=1}^N b_{ij} u_j + \sum_{j=1}^{N^\partial} b_{ij}^\partial u_j^\partial = f_i, \quad 1 \le i \le N,$$

$$u_i^\partial = g_i, \quad 1 \le i \le N^\partial.$$

The matrix form is

$$\tilde{L}_h \tilde{\mathbf{u}} = \tilde{\mathbf{f}}, \tilde{L}_h = \begin{pmatrix} L_h & B^\partial \\ 0 & I \end{pmatrix}, \tilde{\mathbf{u}} = \begin{pmatrix} \mathbf{u} \\ \mathbf{u}^\partial \end{pmatrix}, \tilde{\mathbf{f}} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}.$$

The discrete maximum principle is

$$\mathcal{L}_h(\tilde{\mathbf{u}})_i \le 0, 1 \le i \le N \Longrightarrow \max_i u_i \le \max\{0, \max_i u_i^\partial\}, \tag{3.1}$$

which implies

$$\mathcal{L}_h(\tilde{\mathbf{u}})_i = 0, 1 \le i \le N \Longrightarrow |u_i| \le \max_i |u_i^\partial|.$$

The following result was proven in [6]:

**Theorem 3.1.** *A finite difference operator $\mathcal{L}_h$ satisfies the discrete maximum principle* (3.1) *if $\tilde{L}_h^{-1} \ge 0$ and all row sums of $\tilde{L}_h$ are non-negative.*

With the same $\bar{L}_h$ as defined in the previous section, it suffices to have $\bar{L}_h^{-1} \ge 0$, see [14]:

**Theorem 3.2.** *If $\bar{L}_h^{-1} \ge 0$, then $\tilde{L}_h^{-1} \ge 0$ thus $L_h^{-1} \ge 0$. Moreover, if row sums of $\bar{L}_h$ are non-negative, then the finite difference operator $\mathcal{L}_h$ satisfies the discrete maximum principle.*

Let $\mathbf{1}$ be an abstract vector of the same shape as $\bar{\mathbf{u}}$ with all ones. For the $Q^2$ spectral element method, we have that $(\bar{L}_h \mathbf{1})_{i,j} = 1$ if $(x_i, y_j) \in \partial\Omega$ and $(\bar{L}_h \mathbf{1})_{i,j} = 0$ if $(x_i, y_j) \in \Omega$, which implies the row sums of $\bar{L}_h$ are non-negative. Thus from now on, we only need to discuss the monotonicity of the matrix $\bar{L}_h$.

## 3.2 Lorenz's sufficient condition for monotonicity

**Definition 1.** Let $\mathcal{N} = \{1, 2, \ldots, n\}$. For $\mathcal{N}_1, \mathcal{N}_2 \subset \mathcal{N}$, we say a matrix $A$ of size $n \times n$ connects $\mathcal{N}_1$ with $\mathcal{N}_2$ if

$$\forall i_0 \in \mathcal{N}_1, \exists i_r \in \mathcal{N}_2, \exists i_1, \ldots, i_{r-1} \in \mathcal{N} \quad \text{s.t.} \quad a_{i_{k-1} i_k} \neq 0, \quad k = 1, \cdots, r. \tag{3.2}$$

If perceiving $A$ as a directed graph adjacency matrix of vertices labeled by $\mathcal{N}$, then (3.2) simply means that there exists a directed path from any vertex in $\mathcal{N}_1$ to at least one vertex in $\mathcal{N}_2$. In particular, if $\mathcal{N}_1 = \varnothing$, then any matrix $A$ connects $\mathcal{N}_1$ with $\mathcal{N}_2$.

Given a square matrix $A$ and a column vector $\mathbf{x}$, we define

$$\mathcal{N}^0(A\mathbf{x}) = \{i : (A\mathbf{x})_i = 0\}, \quad \mathcal{N}^+(A\mathbf{x}) = \{i : (A\mathbf{x})_i > 0\}.$$

Given a matrix $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, define its diagonal, off-diagonal, positive and negative off-diagonal parts as $n \times n$ matrices $A_d$, $A_a$, $A_a^+$, $A_a^-$:

$$(A_d)_{ij} = \begin{cases} a_{ii}, & \text{if} \quad i = j \\ 0, & \text{if} \quad i \neq j \end{cases}, \quad A_a = A - A_d,$$

$$(A_a^+)_{ij} = \begin{cases} a_{ij}, & \text{if} \quad a_{ij} > 0, \quad i \neq j \\ 0, & \text{otherwise.} \end{cases}, \quad A_a^- = A_a - A_a^+.$$

The following two results were proven in [19]. See also [14] for a detailed proof.

**Theorem 3.3.** *If $A \leq M_1 M_2 \cdots M_k L$ where $M_1, \cdots, M_k$ are nonsingular M-matrices and $L_a \leq 0$, and there exists a nonzero vector $\mathbf{e} \geq 0$ such that $A\mathbf{e} \geq 0$ and one of the matrices $M_1, \cdots, M_k, L$ connects $\mathcal{N}^0(A\mathbf{e})$ with $\mathcal{N}^+(A\mathbf{e})$. Then $M_k^{-1} M_{k-1}^{-1} \cdots M_1^{-1} A$ is an M-matrix, thus $A$ is a product of $k + 1$ nonsingular M-matrices and $A^{-1} \geq 0$.*

**Theorem 3.4** (Lorenz's condition). *If $A_a^-$ has a decomposition: $A_a^- = A^z + A^s = (a_{ij}^z) + (a_{ij}^s)$ with $A^s \leq 0$ and $A^z \leq 0$, such that*

$$A_d + A^z \text{ is a nonsingular M-matrix}, \tag{3.3a}$$

$$A_a^+ \leq A^z A_d^{-1} A^s \text{ or equivalently } \forall a_{ij} > 0 \text{ with } i \neq j, a_{ij} \leq \sum_{k=1}^n a_{ik}^z a_{kk}^{-1} a_{kj}^s, \tag{3.3b}$$

$$\exists \mathbf{e} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbf{e} \geq 0 \text{ with } A\mathbf{e} \geq 0 \text{ s.t. } A^z \text{ or } A^s \text{ connects } \mathcal{N}^0(A\mathbf{e}) \text{ with } \mathcal{N}^+(A\mathbf{e}). \tag{3.3c}$$

*Then $A$ is a product of two nonsingular M-matrices thus $A^{-1} \geq 0$.*

**Proposition 1.** The matrix $L$ in Theorem 3.3 must be an M-matrix.

*Proof.* Let $M^{-1} = M_k^{-1} M_{k-1}^{-1} ... M_1^{-1}$, following the proof of Theorem 7 in [14], then $M^{-1} A \mathbf{e} \geq c A \mathbf{e}$ for some positive number $c$. Then $A \mathbf{e} \geq 0 \Rightarrow M^{-1} A \mathbf{e} \geq 0$. Now since $\mathbf{e} \geq 0$, $M^{-1} A \leq L \Rightarrow 0 \leq (L - M^{-1} A) \mathbf{e} \Rightarrow M^{-1} A \mathbf{e} \leq L \mathbf{e}$ thus $L \mathbf{e} \geq 0$.

Assume $L$ connects $\mathcal{N}^0(A\mathbf{e})$ with $\mathcal{N}^+(A\mathbf{e})$. Since $M^{-1} A \mathbf{e} \leq L \mathbf{e}$, $\mathcal{N}^0(L\mathbf{e}) \subseteq \mathcal{N}^0(A\mathbf{e})$ and $\mathcal{N}^+(A\mathbf{e}) \subseteq \mathcal{N}^+(L\mathbf{e})$, so $L$ also connects $\mathcal{N}^0(L\mathbf{e})$ with $\mathcal{N}^+(L\mathbf{e})$.

Assume $M_i$ connects $\mathcal{N}^0(A\mathbf{e})$ with $\mathcal{N}^+(A\mathbf{e})$, following the proof of Theorem 7 in [14], we have $M^{-1} A \mathbf{e} > 0$. Now $L$ trivially connects $\mathcal{N}^0(L\mathbf{e})$ with $\mathcal{N}^+(L\mathbf{e})$ since $L \mathbf{e} \geq M^{-1} A \mathbf{e} \Rightarrow L \mathbf{e} > 0$ and $\mathcal{N}^0(L\mathbf{e}) = \emptyset$.

Then Theorem 6 in [14] applies to show $L$ is an M-matrix. $\qquad\square$

In practice, the condition (3.3c) can be difficult to verify. For variational difference schemes, the vector $\mathbf{e}$ can be taken as $\mathbf{1}$ consisting of all ones, then the condition (3.3c) can be simplified. The following theorem was proven in [14].

**Theorem 3.5.** *Let $A$ denote the matrix representation of the variational difference scheme* (2.3) *with $Q^2$ basis solving $-\nabla \cdot (a\nabla) u + cu = f$. Assume $A_a^-$ has a decomposition $A_a^- = A^z + A^s$ with $A^s \leq 0$ and $A^z \leq 0$. Then $A^{-1} \geq 0$ if the following are satisfied:*

1. $(A_d + A^z)\mathbf{1} \neq \mathbf{0}$ and $(A_d + A^z)\mathbf{1} \geq 0$;

2. $A_a^+ \leq A^z A_d^{-1} A^s$;

3. *For $c(x,y) \geq 0$, either $A^z$ or $A^s$ has the same sparsity pattern as $A_a^-$. If $c(x,y) > 0$, then this condition can be removed.*

## 3.3 A relaxed Lorenz's condition

In practice, both (3.3a) and (3.3b) impose mesh constraints for the $Q^2$ spectral element method on non-uniform meshes. The condition (3.3a) can be relaxed as the following:

**Theorem 3.6** (A relaxed Lorenz's condition). *If $A_a^-$ has a decomposition: $A_a^- = A^z + A^s = (a_{ij}^z) + (a_{ij}^s)$ with $A^s \leq 0$ and $A^z \leq 0$, and there exists a diagonal matrix $A_{d*} \geq A_d$ such that*

$$A_d^* + A^z \text{ is a nonsingular M-matrix,} \tag{3.4a}$$
$$A_a^+ \leq A^z A_{d*}^{-1} A^s, \tag{3.4b}$$
$$\exists \mathbf{e} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbf{e} \geq 0 \text{ with } A\mathbf{e} \geq 0 \text{ s.t. } A^z \text{ or } A^s \text{ connects } \mathcal{N}^0(A\mathbf{e}) \text{ with } \mathcal{N}^+(A\mathbf{e}). \tag{3.4c}$$

*Then $A$ is a product of two nonsingular M-matrices thus $A^{-1} \geq 0$.*

*Proof.* It is straightforward that $A = A_d + A_a^+ + A^z + A^s \leq A_{d*} + A^z + A^s + A^z A_{d*}^{-1} A^s = (A_{d*} + A^z)(I + A_{d*}^{-1} A^s)$. By (3.4c), either $A_{d*} + A^z$ or $I + A_{d*}^{-1} A^s$ connects $\mathcal{N}^0(A\mathbf{e})$ with $\mathcal{N}^+(A\mathbf{e})$. By applying Theorem 3.3 for the case $k = 1$, $M_1 = A_{d*} + A^z$ and $L = I + A_{d*}^{-1} A^s$, we get $A^{-1} \geq 0$. $\qquad\square$

**Remark 1.** Since $A_d \leq A_{d^*}$, only (3.4a) is more relaxed than (3.3a), and (3.4b) is more stringent than (3.3b). However, we will show in next section that it is possible to construct $A_{d^*}$ such that (3.3b) and (3.4b) impose identical mesh constraints.

With Theorem A.1, combining Theorem 3.6 and Theorem 3.5, we have:

**Theorem 3.7.** *Let $A$ denote the matrix representation of the variational difference scheme* (2.3) *with $Q^2$ basis solving $-\nabla \cdot (a\nabla)u + cu = f$. Assume $A_a^-$ has a decomposition $A_a^- = A^z + A^s$ with $A^s \leq 0$ and $A^z \leq 0$ and there exists a diagonal matrix $A_{d^*} \geq A_d$. Then $A^{-1} \geq 0$ if the following are satisfied:*

1. *$(A_{d^*} + A^z)\mathbf{1} \neq \mathbf{0}$ and $(A_{d^*} + A^z)\mathbf{1} \geq 0$;*

2. *$A_a^+ \leq A^z A_{d^*}^{-1} A^s$;*

3. *For $c(x,y) \geq 0$, either $A^z$ or $A^s$ has the same sparsity pattern as $A_a^-$. If $c(x,y) > 0$, then this condition can be removed.*

# 4 Monotonicity of $Q^2$ spectral element method on quasi-uniform meshes

The $Q^2$ spectral element method has been proven monotone on a uniform mesh for Laplacian operator without any mesh constraints [14]. In this section, we will discuss its monotonicity for the Laplacian operator on quasi-uniform meshes. The discussion in this section can be easily extended to more general cases such as $\mathcal{L}u = -\Delta u + cu$ and Neumann boundary conditions. For simplicity, we only discuss the Laplacian case $\mathcal{L}u = -\Delta u$ and Dirichlet boundary conditions.

Consider a grid $(x_i, y_j)$ $(i,j = 0,1,\ldots,n+1)$ for a rectangular domain $[0,1] \times [0,1]$ where $n$ must be odd and $i,j = 0, n+1$ correspond to boundary points. Let $u_{ij}$ denote the numerical solution at $(x_i, y_j)$. Let $\bar{\mathbf{u}}$ denote an abstract vector consisting of $u_{ij}$ for $i,j = 0,1,2,\cdots,n,n+1$. Let $\bar{\mathbf{f}}$ denote an abstract vector consisting of $f_{ij}$ for $i,j = 1,2,\cdots,n$ and the boundary condition $g$ at the boundary grid points. Then the matrix vector representation of (2.3) with $Q^2$ basis is $\bar{L}_h \bar{\mathbf{u}} = \bar{\mathbf{f}}$.

The focus of this section is to show $\bar{L}_h^{-1} \geq 0$ under suitable mesh constraints for quasi-uniform meshes. Moreover, it is straightforward to verify that $(\bar{L}_h \mathbf{1})_{i,j} = 0$ for interior points $(x_i, y_j)$ and $(\bar{L}_h \mathbf{1})_{i,j} = 1$ for boundary points $(x_i, y_j)$. Thus by Section 3.1, the scheme also satisfies the discrete maximum principle.

For simplicity, in the rest of this section we use $A$ to denote the matrix $\bar{L}_h$ and let $\mathcal{A}$ be the linear operator corresponding to the matrix $A$. For convenience, we can also regard the abstract vector $\bar{\mathbf{u}}$ as a matrix of size $(n+2) \times (n+2)$. Then by our notation, the mapping $\mathcal{A} : \mathbb{R}^{(n+2) \times (n+2)} \to \mathbb{R}^{(n+2) \times (n+2)}$ is given as $\mathcal{A}(\bar{\mathbf{u}})_{i,j} := (\bar{L}_h \bar{\mathbf{u}})_{i,j}$.

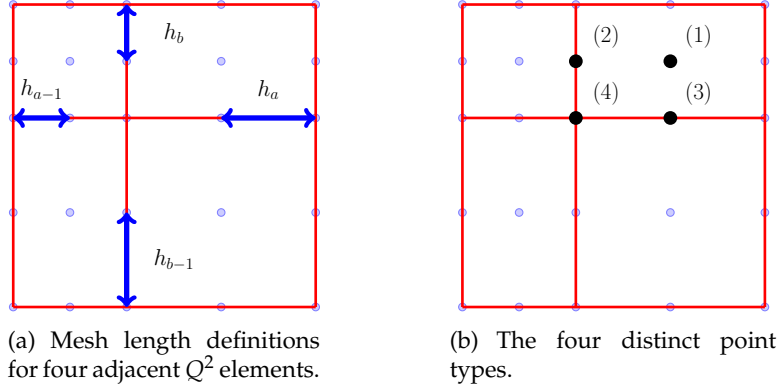(a) Mesh length definitions for four adjacent $Q^2$ elements.

(b) The four distinct point types.

Figure 2: A non-uniform mesh for $Q^2$ spectral element method. Each edge in a cell has length $2h$.

## 4.1 The scheme in two dimensions

For boundary points $(x_i, y_j) \in \partial\Omega$, the scheme is $\mathcal{A}(\bar{\mathbf{u}})_{i,j} := u_{i,j} = g_{i,j}$. The scheme for interior grid points $(x_i, y_j) \in \Omega$ on a non-uniform mesh can be given on four distinct types of points shown in Figure 2 (b). For simplicity, from now on, we will use *edge center (2)* to denote an interior edge center for an edge parallel to the y-axis, and *edge center (3)* to denote an interior edge center for an edge parallel to the x-axis. The scheme at an interior grid point

is given as $\mathcal{A}(\bar{\mathbf{u}})_{i,j} = f_{i,j}$ with

$$\mathcal{A}(\bar{\mathbf{u}})_{i,j} := \frac{2h_a^2 + 2h_b^2}{h_a^2 h_b^2} u_{i,j} - \left( \frac{1}{h_a^2} u_{i+1,j} + \frac{1}{h_a^2} u_{i-1,j} + \frac{1}{h_b^2} u_{i,j+1} + \frac{1}{h_b^2} u_{i,j-1} \right) \tag{4.1}$$

if $(x_i, y_j)$ is a cell center;

$$\mathcal{A}(\bar{\mathbf{u}})_{i,j} := \frac{7h_b^2 + 4h_a h_{a-1}}{2h_a h_{a-1} h_b^2} u_{i,j} - \frac{4}{h_a(h_a + h_{a-1})} u_{i+1,j} - \frac{4}{h_{a-1}(h_a + h_{a-1})} u_{i-1,j}$$
$$- \frac{1}{h_b^2} u_{i,j+1} - \frac{1}{h_b^2} u_{i,j-1} + \frac{1}{2h_a(h_a + h_{a-1})} u_{i+2,j} + \frac{1}{2h_{a-1}(h_a + h_{a-1})} u_{i-2,j},$$

if $(x_i, y_j)$ is edge center (2);

$$\mathcal{A}(\bar{\mathbf{u}})_{i,j} := \frac{7h_a^2 + 4h_b h_{b-1}}{2h_b h_{b-1} h_a^2} u_{i,j} - \frac{4}{h_b(h_b + h_{b-1})} u_{i,j+1} - \frac{4}{h_{b-1}(h_b + h_{b-1})} u_{i,j-1}$$
$$- \frac{1}{h_a^2} u_{i+1,j} - \frac{1}{h_a^2} u_{i-1,j} + \frac{1}{2h_b(h_b + h_{b-1})} u_{i,j+2} + \frac{1}{2h_{b-1}(h_b + h_{b-1})} u_{i,j-2},$$

if $(x_i, y_j)$ is edge center (3);

$$\mathcal{A}(\bar{\mathbf{u}})_{i,j} := \frac{7h_a h_{a-1} + 7h_b h_{b-1}}{2h_a h_{a-1} h_b h_{b-1}} u_{i,j} - \left[ \frac{4}{h_a(h_a + h_{a-1})} u_{i+1,j} + \frac{4}{h_{a-1}(h_a + h_{a-1})} u_{i-1,j} \right.$$
$$\left. + \frac{4}{h_b(h_b + h_{b-1})} u_{i,j+1} + \frac{4}{h_{b-1}(h_b + h_{b-1})} u_{i,j-1} \right] + \frac{1}{2h_a(h_a + h_{a-1})} u_{i+2,j}$$
$$+ \frac{1}{2h_{a-1}(h_a + h_{a-1})} u_{i-2,j} + \frac{1}{2h_b(h_b + h_{b-1})} u_{i,j+2} + \frac{1}{2h_{b-1}(h_b + h_{b-1})} u_{i,j-2},$$

if $(x_i, y_j)$ is an interior knot.

For a uniform mesh $h_a = h_{a-1} = h_b = h_{b-1} = h$, the scheme reduces to (2.4).

## 4.2   The Decomposition of $A_a^-$

Next, by the same notations defined in Section 3.2, we will decompose the matrix $A = A_d + A_a^- + A_a^+$ and $A_a^- = A^z + A^s$ to verify Theorem 3.5. We will use $\mathcal{A}_a^-$, $\mathcal{A}_a^+$, $\mathcal{A}^z$ and $\mathcal{A}^s$ to

denote linear operators for corresponding matrices. First, for the diagonal part we have

$$\mathcal{A}_d(\bar{\mathbf{u}})_{i,j} = u_{i,j}, \quad \text{if } (x_i, y_j) \text{ is a boundary point;}$$

$$\mathcal{A}_d(\bar{\mathbf{u}})_{i,j} = \frac{2h_a^2 + 2h_b^2}{h_a^2 h_b^2} u_{i,j}, \quad \text{if } (x_i, y_j) \text{ is a cell center;}$$

$$\mathcal{A}_d(\bar{\mathbf{u}})_{i,j} = \frac{7h_b^2 + 4h_a h_{a-1}}{2h_a h_{a-1} h_b^2} u_{i,j}, \quad \text{if } (x_i, y_j) \text{ is edge center (2);}$$

$$\mathcal{A}_d(\bar{\mathbf{u}})_{i,j} = \frac{7h_a^2 + 4h_b h_{b-1}}{2h_b h_{b-1} h_a^2} u_{i,j}, \quad \text{if } (x_i, y_j) \text{ is edge center (3);}$$

$$\mathcal{A}_d(\bar{\mathbf{u}})_{i,j} = \frac{7h_b h_{b-1} + 7h_a h_{a-1}}{2h_a h_{a-1} h_b h_{b-1}} u_{i,j}, \quad \text{if } (x_i, y_j) \text{ is an interior knot.}$$

Notice that for a boundary point $(x_i, y_j) \in \partial\Omega$ we have $\mathcal{A}(\bar{\mathbf{u}})_{i,j} = \mathcal{A}_d(\bar{\mathbf{u}})_{i,j} = u_{i,j}$, thus for off-diagonal parts, we only need to look at the interior grid points. For positive off-diagonal entries, we have

$$\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j} = 0, \quad \text{if } (x_i, y_j) \text{ is a cell center;}$$

$$\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j} = \frac{1}{2h_a(h_a + h_{a-1})} u_{i+2,j} + \frac{1}{2h_{a-1}(h_a + h_{a-1})} u_{i-2,j}, \quad \text{edge center (2);}$$

$$\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j} = \frac{1}{2h_b(h_b + h_{b-1})} u_{i,j+2} + \frac{1}{2h_{b-1}(h_b + h_{b-1})} u_{i,j-2}, \quad \text{edge center (3);}$$

$$\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j} = \frac{1}{2h_a(h_a + h_{a-1})} u_{i+2,j} + \frac{1}{2h_{a-1}(h_a + h_{a-1})} u_{i-2,j} + \frac{1}{2h_b(h_b + h_{b-1})} u_{i,j+2}$$

$$+ \frac{1}{2h_{b-1}(h_b + h_{b-1})} u_{i,j-2}, \quad \text{if } (x_i, y_j) \text{ is an interior knot.}$$

Then we perform a decomposition $A_a^- = A^z + A^s$, which depends on two constants

$0 < \epsilon_1 \leq 1$ and $0 < \epsilon_2 \leq 1$.

$$\mathcal{A}^z(\bar{\mathbf{u}})_{i,j} = -\epsilon_1 \left( \frac{1}{h_a^2} u_{i+1,j} + \frac{1}{h_a^2} u_{i-1,j} + \frac{1}{h_b^2} u_{i,j+1} + \frac{1}{h_b^2} u_{i,j-1} \right), \quad \text{if } (x_i, y_j) \text{ is a cell center;}$$

$$\mathcal{A}^z(\bar{\mathbf{u}})_{i,j} = -\epsilon_1 \left( \frac{1}{h_b^2} u_{i,j+1} + \frac{1}{h_b^2} u_{i,j-1} \right) - \epsilon_2 \left[ \frac{4}{h_a(h_a + h_{a-1})} u_{i+1,j} + \frac{4}{h_{a-1}(h_a + h_{a-1})} u_{i-1,j} \right],$$
$$\text{if } (x_i, y_j) \text{ is edge center (2);}$$

$$\mathcal{A}^z(\bar{\mathbf{u}})_{i,j} = -\epsilon_1 \left( \frac{1}{h_a^2} u_{i+1,j} + \frac{1}{h_a^2} u_{i-1,j} \right) - \epsilon_2 \left[ \frac{4}{h_b(h_b + h_{b-1})} u_{i,j+1} + \frac{4}{h_{b-1}(h_b + h_{b-1})} u_{i,j-1} \right],$$
$$\text{if } (x_i, y_j) \text{ is edge center (3);}$$

$$\mathcal{A}^z(\bar{\mathbf{u}})_{i,j} = -\epsilon_2 \left[ \frac{4}{h_a(h_a + h_{a-1})} u_{i+1,j} + \frac{4}{h_{a-1}(h_a + h_{a-1})} u_{i-1,j} \right.$$
$$\left. + \frac{4}{h_b(h_b + h_{b-1})} u_{i,j+1} + \frac{4}{h_{b-1}(h_b + h_{b-1})} u_{i,j-1} \right], \quad \text{if } (x_i, y_j) \text{ is an interior knot.}$$

Notice that $A^z$ defined above has exactly the same sparsity pattern as $A_a^-$ for $0 < \epsilon_1 \leq 1$ and $0 < \epsilon_2 \leq 1$. Let $A^s = A_a^- - A^z$ then $A^s \leq 0$.

## 4.3 Mesh constraints for $A^z A_d^{-1} A^s \geq A_a^+$

In order to verify $A^z A_d^{-1} A^s \geq A_a^+$, we only need to discuss nonzero entries in the output of $\mathcal{A}_a^+(\bar{\mathbf{u}})$ since $A^z A_d^{-1} A^s \geq 0$.
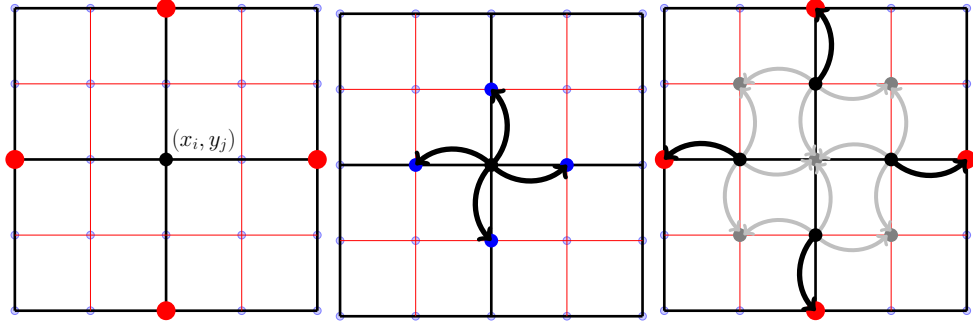


(a) Four red dots denote non-zero entry locations in $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}$

(b) Stencil of $\mathcal{A}^z(\bar{\mathbf{u}})_{i,j}$.

(c) Stencil of $\mathcal{A}^z A_d^{-1} \mathcal{A}^s(\bar{\mathbf{u}})_{i,j}$.

Figure 3: Stencil of operators at an interior knot $(x_i, y_j)$. The four red dots are the locations/entries where $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}$ are nonzero. Gray nodes in (c) represent positive entries that can be discarded for the purposes of verifying (3.4b). The mesh is illustrated as a uniform one only for simplicity.

First consider the case that $(x_i, y_j)$ is an interior knot. Figure 3 (a) shows the positive coefficients in the output of $\mathcal{A}_a^+(\bar{\mathbf{u}})_{ij}$ at a knot $(x_i, y_j)$. Figure 3 (b) shows the stencil of

$\mathcal{A}^z(\bar{\mathbf{u}})_{ij}$. Thus $\mathcal{A}^z(\bar{\mathbf{u}})$ acting as an operator on $[\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})$ at a knot is:

$$[\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j} = -4\epsilon_2\left[\frac{1}{h_a(h_{a-1}+h_a)}[\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i+1,j} + \frac{1}{h_{a-1}(h_{a-1}+h_a)}[\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i-1,j}\right.$$
$$\left. + \frac{1}{h_b(h_{b-1}+h_b)}[\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j+1} + \frac{1}{h_{b-1}(h_{b-1}+h_b)}[\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j-1}\right].$$

In the expression above, the output of the operator $\mathcal{A}^z(\bar{\mathbf{u}})_{ij}$ are at interior edge centers as shown in Figure 3 (b). Hence $[\mathcal{A}_d^{-1}\mathcal{A}^s]$ will act on these edge centers with the mesh lengths corresponding to Figure 2. Carefully considering the mesh lengths and operations of $\mathcal{A}_d^{-1}$ at these points gives:

$$[\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j} = -4\epsilon_2\left[\frac{1}{h_a(h_{a-1}+h_a)}\frac{2h_bh_{b-1}h_a^2}{7h_a^2+4h_bh_{b-1}}\mathcal{A}^s(\bar{\mathbf{u}})_{i+1,j}\right.$$
$$+\frac{1}{h_{a-1}(h_{a-1}+h_a)}\frac{2h_bh_{b-1}h_{a-1}^2}{7h_{a-1}^2+4h_bh_{b-1}}\mathcal{A}^s(\bar{\mathbf{u}})_{i-1,j} + \frac{1}{h_b(h_{b-1}+h_b)}\frac{2h_ah_{a-1}h_b^2}{7h_b^2+4h_ah_{a-1}}\mathcal{A}^s(\bar{\mathbf{u}})_{i,j+1}$$
$$\left.+\frac{1}{h_{b-1}(h_{b-1}+h_b)}\frac{2h_ah_{a-1}h_{b-1}^2}{7h_{b-1}^2+4h_ah_{a-1}}\mathcal{A}^s(\bar{\mathbf{u}})_{i,j-1}\right], \quad \text{if } (x_i,y_j) \text{ is an interior knot.}$$

Next consider the effect of $\mathcal{A}^s(\bar{\mathbf{u}})$ operator which has the same sparsity pattern as $\mathcal{A}^z(\bar{\mathbf{u}})$. Figure 3 (c) shows the stencil of $[\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j}$ for an interior knot. Recall that $A^z \leq 0$, $A^s \leq 0$, and $A_d^{-1} \geq 0$, thus we have $A^z A_d^{-1} A^s \geq 0$. So we only need to compare the outputs of $[\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j}$ and $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}$ at nonzero entries of $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}$, i.e., the four red dots in Figure 3 (a) and Figure 3 (c).

Thus we only need coefficients of $u_{i+2,j}, u_{i-2,j}, u_{i,j+2}$, and $u_{i,j-2}$ in the final expression of $[\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s](\mathbf{u})_{i,j}$, which are found to be

$$u_{i+2,j}: \quad 4\epsilon_2(1-\epsilon_1)\frac{1}{h_a(h_{a-1}+h_a)}\frac{2h_bh_{b-1}h_a^2}{7h_a^2+4h_bh_{b-1}}\frac{1}{h_a^2}$$

$$u_{i-2,j}: \quad 4\epsilon_2(1-\epsilon_1)\frac{1}{h_{a-1}(h_{a-1}+h_a)}\frac{2h_bh_{b-1}h_{a-1}^2}{7h_{a-1}^2+4h_bh_{b-1}}\frac{1}{h_{a-1}^2}$$

$$u_{i,j+2}: \quad 4\epsilon_2(1-\epsilon_1)\frac{1}{h_b(h_{b-1}+h_b)}\frac{2h_ah_{a-1}h_b^2}{7h_b^2+4h_ah_{a-1}}\frac{1}{h_b^2}$$

$$u_{i,j-2}: \quad 4\epsilon_2(1-\epsilon_1)\frac{1}{h_{b-1}(h_{b-1}+h_b)}\frac{2h_ah_{a-1}h_{b-1}^2}{7h_{b-1}^2+4h_ah_{a-1}}\frac{1}{h_{b-1}^2}$$

In order to maintain $A_a^+ \leq A^z A_d^{-1} A^s$, by comparing to the coefficients of $u_{i+2,j}$ for $\mathcal{A}_a^+(\bar{\mathbf{u}})$, we obtain a mesh constraint $4\epsilon_2(1-\epsilon_1)\frac{2h_bh_{b-1}}{7h_a^2+4h_bh_{b-1}} \geq \frac{1}{2}$. Similar constraints are obtained by comparing other coefficients at $u_{i,j\mp2}$ and $u_{i-2,j}$. Define

$$\ell(\epsilon_1,\epsilon_2) = 4\epsilon_2(1-\epsilon_1).$$

Then the following constraints are sufficient for $\mathcal{A}_a^+(\bar{\mathbf{u}})$ to be controlled by $\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s(\bar{\mathbf{u}})$ at an interior knot:

$$h_a h_{a-1} \geq \frac{7}{4\ell-4}\max\{h_b^2, h_{b-1}^2\}, \quad h_b h_{b-1} \geq \frac{7}{4\ell-4}\max\{h_a^2, h_{a-1}^2\}. \qquad (4.2a)$$

Second, we need to discuss the case when $(x_i, y_j)$ is an interior edge center. Without loss of generality, assume $(x_i, y_j)$ is an interior edge center of an edge parallel to the y-axis. Then similar to the interior knot case, the output coefficients of $[\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s](\bar{\mathbf{u}})_{i,j}$ at the relevant non-zero entries of $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}$ are:

$$u_{i+2,j}: \quad 4\epsilon_2(1-\epsilon_1)\frac{1}{h_a(h_{a-1}+h_a)}\frac{h_a^2 h_b^2}{2h_a^2+2h_b^2}\frac{1}{h_a^2}$$

$$u_{i-2,j}: \quad 4\epsilon_2(1-\epsilon_1)\frac{1}{h_{a-1}(h_{a-1}+h_a)}\frac{h_{a-1}^2 h_b^2}{2h_{a-1}^2+2h_b^2}\frac{1}{h_{a-1}^2}$$

By comparing with coefficients of $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}$, we get $\frac{h_b^2}{h_a^2+h_b^2} \geq \frac{1}{\ell}$, $\frac{h_b^2}{h_{a-1}^2+h_b^2} \geq \frac{1}{\ell}$. To ensure $\mathcal{A}_a^+(\bar{\mathbf{u}})$ is controlled by $\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s(\bar{\mathbf{u}})$ at edge centers, it suffices to have:

$$min\{h_a, h_{a-1}\} \geq \sqrt{\frac{1}{\ell-1}}max\{h_b, h_{b-1}\}, \quad min\{h_b, h_{b-1}\} \geq \sqrt{\frac{1}{\ell-1}}max\{h_a, h_{a-1}\}. \quad (4.2b)$$

Note that $\mathcal{A}_a^+(\bar{\mathbf{u}})_{i,j}=0$ if $(x_i, y_j)$ is a cell center. Since $\mathcal{A}^z\mathcal{A}_d^{-1}\mathcal{A}^s(\bar{\mathbf{u}}) \geq 0$, there is no mesh constraint to enforce the inequality at cell centers.

## 4.4 Mesh constraints for $A_d + A^z$ being an M-matrix

Let $\mathcal{B} = \mathcal{A}_d + \mathcal{A}^z$. Then $\mathcal{B}(\mathbf{1})_{i,j} = 1$ for a boundary point $(x_i, y_j)$. For interior points, we have:

$$\mathcal{B}(\mathbf{1})_{i,j} = -\epsilon_1\left(\frac{1}{h_a^2}+\frac{1}{h_a^2}+\frac{1}{h_b^2}+\frac{1}{h_b^2}\right) + \frac{2h_a^2+2h_b^2}{h_a^2 h_b^2} = (1-\epsilon_1)\frac{2h_a^2+2h_b^2}{h_a^2 h_b^2}, \quad \text{cell center;}$$

$$\mathcal{B}(\mathbf{1})_{i,j} = -\epsilon_1\left(\frac{1}{h_b^2}+\frac{1}{h_b^2}\right) - \epsilon_2\left[\frac{4}{h_a(h_a+h_{a-1})}+\frac{4}{h_{a-1}(h_a+h_{a-1})}\right] + \frac{7h_b^2+4h_a h_{a-1}}{2h_a h_{a-1}h_b^2}$$

$$= (1-\epsilon_1)\frac{2}{h_b^2}+(1-\frac{8}{7}\epsilon_2)\frac{7}{2h_a h_{a-1}}, \quad \text{edge center (2);}$$

$$\mathcal{B}(\mathbf{1})_{i,j} = -\epsilon_1\left(\frac{1}{h_a^2}+\frac{1}{h_a^2}\right) - \epsilon_2\left[\frac{4}{h_b(h_b+h_{b-1})}+\frac{4}{h_{b-1}(h_b+h_{b-1})}\right] + \frac{7h_a^2+4h_b h_{b-1}}{2h_b h_{b-1}h_a^2}$$

$$= (1-\epsilon_1)\frac{2}{h_a^2}+(1-\frac{8}{7}\epsilon_2)\frac{7}{2h_b h_{b-1}}, \quad \text{edge center (3);}$$

$$\mathcal{B}(\mathbf{1})_{i,j} = -\epsilon_2\left[\frac{4}{h_a(h_a+h_{a-1})}+\frac{4}{h_{a-1}(h_a+h_{a-1})}+\frac{4}{h_b(h_b+h_{b-1})}+\frac{4}{h_{b-1}(h_b+h_{b-1})}\right]$$

$$+ \frac{7h_b h_{b-1}+7h_a h_{a-1}}{2h_a h_{a-1}h_b h_{b-1}} = (1-\frac{8}{7}\epsilon_2)\frac{7h_b h_{b-1}+7h_a h_{a-1}}{2h_a h_{a-1}h_b h_{b-1}}, \quad \text{interior knot.}$$

Notice that larger values of $\ell$ give better mesh constraints in (4.2). And we have $\sup_{0<\epsilon_1,\epsilon_2\leq 1}\ell(\epsilon_1,\epsilon_2)=\sup_{0<\epsilon_1,\epsilon_2\leq 1}4\epsilon_2(1-\epsilon_1)=4$. In order to apply Theorem A.1 for $A_d+A^z$ be an M-matrix, we need $[\mathcal{A}_d+\mathcal{A}^z](\mathbf{1})\geq 0$. This is true if and only if $\epsilon_1\leq 1$ and $\epsilon_2\leq\frac{7}{8}$, which only give $\sup_{0<\epsilon_1\leq 1,0<\epsilon_2\leq\frac{7}{8}}\ell(\epsilon_1,\epsilon_2)=3.5$.

## 4.5   Improved mesh constraints by the relaxed Lorenz's condition

To get a better mesh constraint, the constraint on $\epsilon_2$ can be relaxed so that the value of $\ell(\epsilon_1,\epsilon_2)$ can be improved. One observation from Section 4.3 is that the value of $\mathcal{A}_d(\bar{\mathbf{u}})_{i,j}$ for $(x_i,y_j)$ being a knot is not used for verifying $A_a^+\leq A^z A_d^{-1}A^s$ (for both interior knots and edge centers). To this end, we define a new diagonal matrix $A_{d^*}$, which is different from $A_d$ only at the interior knots.

$$\mathcal{A}_{d^*}(\bar{\mathbf{u}})_{i,j}=u_{i,j}=\mathcal{A}_d(\bar{\mathbf{u}})_{i,j},\quad \text{if } (x_i,y_j) \text{ is a boundary point;}$$

$$\mathcal{A}_{d^*}(\bar{\mathbf{u}})_{i,j}=\frac{2h_a^2+2h_b^2}{h_a^2 h_b^2}u_{i,j}=\mathcal{A}_d(\bar{\mathbf{u}})_{i,j},\quad \text{if } (x_i,y_j) \text{ is a cell center;}$$

$$\mathcal{A}_{d^*}(\bar{\mathbf{u}})_{i,j}=\frac{7h_b^2+4h_a h_{a-1}}{2h_a h_{a-1} h_b^2}u_{i,j}=\mathcal{A}_d(\bar{\mathbf{u}})_{i,j},\quad \text{edge center (2);}$$

$$\mathcal{A}_{d^*}(\bar{\mathbf{u}})_{i,j}=\frac{7h_a^2+4h_b h_{b-1}}{2h_b h_{b-1} h_a^2}u_{i,j}=\mathcal{A}_d(\bar{\mathbf{u}})_{i,j},\quad \text{edge center (3);}$$

$$\mathcal{A}_{d^*}(\bar{\mathbf{u}})_{i,j}=\frac{8h_b h_{b-1}+8h_a h_{a-1}}{2h_a h_{a-1} h_b h_{b-1}}u_{i,j}\neq\mathcal{A}_d(\bar{\mathbf{u}})_{i,j},\quad \text{if } (x_i,y_j) \text{ is an interior knot.}$$

Since the values of $\mathcal{A}_d(\bar{\mathbf{u}})_{i,j}$ for $(x_i,y_j)$ being a knot is not involved in Section 4.3, the same discussion in Section 4.3 also holds for verifying $A_a^+\leq A^z A_{d^*}^{-1}A^s$. Namely, under mesh constraints (4.2), we also have $A_a^+\leq A^z A_{d^*}^{-1}A^s$.

Let $B^* = A_{d^*} + A^z$, then the row sums of $B^*$ are:

$$\mathcal{B}^*(\mathbf{1})_{i,j} = 1, \quad \text{if } (x_i, y_j) \text{ is a boundary point;}$$

$$\mathcal{B}^*(\mathbf{1})_{i,j} = -\epsilon_1 \left( \frac{1}{h_a^2} + \frac{1}{h_a^2} + \frac{1}{h_b^2} + \frac{1}{h_b^2} \right) + \frac{2h_a^2 + 2h_b^2}{h_a^2 h_b^2} = (1-\epsilon_1) \frac{2h_a^2 + 2h_b^2}{h_a^2 h_b^2}, \text{cell center;}$$

$$\mathcal{B}^*(\mathbf{1})_{i,j} = -\epsilon_1 \left( \frac{1}{h_b^2} + \frac{1}{h_b^2} \right) - \epsilon_2 \left[ \frac{4}{h_a(h_a + h_{a-1})} + \frac{4}{h_{a-1}(h_a + h_{a-1})} \right] + \frac{7h_b^2 + 4h_a h_{a-1}}{2h_a h_{a-1} h_b^2}$$

$$= (1-\epsilon_1) \frac{2}{h_b^2} + (1 - \frac{8}{7}\epsilon_2) \frac{7}{2h_a h_{a-1}}, \quad \text{edge center (2);}$$

$$\mathcal{B}^*(\mathbf{1})_{i,j} = -\epsilon_1 \left( \frac{1}{h_a^2} + \frac{1}{h_a^2} \right) - \epsilon_2 \left[ \frac{4}{h_b(h_b + h_{b-1})} + \frac{4}{h_{b-1}(h_b + h_{b-1})} \right] + \frac{7h_a^2 + 4h_b h_{b-1}}{2h_b h_{b-1} h_a^2}$$

$$= (1-\epsilon_1) \frac{2}{h_a^2} + (1 - \frac{8}{7}\epsilon_2) \frac{7}{2h_b h_{b-1}}, \quad \text{edge center (3);}$$

$$\mathcal{B}^*(\mathbf{1})_{i,j} = -\epsilon_2 \left[ \frac{4}{h_a(h_a + h_{a-1})} + \frac{4}{h_{a-1}(h_a + h_{a-1})} + \frac{4}{h_b(h_b + h_{b-1})} + \frac{4}{h_{b-1}(h_b + h_{b-1})} \right]$$

$$+ \frac{8h_b h_{b-1} + 8h_a h_{a-1}}{2h_a h_{a-1} h_b h_{b-1}} = (1-\epsilon_2) \frac{8h_b h_{b-1} + 8h_a h_{a-1}}{2h_a h_{a-1} h_b h_{b-1}}, \quad \text{interior knot.}$$

Now $[\mathcal{A}_{d^*} + \mathcal{A}^z](\mathbf{1})_{i,j} \geq 0$ at cell centers and knots is true if and only if $\epsilon_1 \leq 1$ and $\epsilon_2 \leq 1$.

Next, we will show that the mesh constraints (4.2) with $0 < \epsilon_1 \leq \frac{1}{2}$ and $\epsilon_2 = 1$ are sufficient to ensure $[\mathcal{A}_{d^*} + \mathcal{A}^z](\mathbf{1})_{i,j} \geq 0$ at edge centers. We have $0 < \epsilon_1 \leq \frac{1}{2}, \epsilon_2 = 1 \implies 2 \leq \ell < 4 \implies \frac{7}{4\ell-4} \geq \frac{1}{\ell}$. The mesh constraints (4.2) imply that $h_a h_{a-1} \geq \frac{7}{4\ell-4} h_b^2 \geq \frac{1}{\ell} h_b^2$, thus

$$(1-\epsilon_1) \frac{2}{h_b^2} + (1 - \frac{8}{7}\epsilon_2) \frac{7}{2h_a h_{a-1}} = (1-\epsilon_1) \frac{2}{h_b^2} - \frac{1}{2} \frac{1}{h_a h_{a-1}} = \frac{1}{2} \left[ \frac{\ell}{h_b^2} - \frac{1}{h_a h_{a-1}} \right] \geq 0.$$

Similarly, $(1-\epsilon_1) \frac{2}{h_a^2} + (1 - \frac{8}{7}\epsilon_2) \frac{7}{2h_b h_{b-1}} \geq 0$ also holds.

Therefore, for constants $0 < \epsilon_1 \leq \frac{1}{2}$ and $\epsilon_2 = 1$, we have $[\mathcal{A}_{d^*} + \mathcal{A}^z](\mathbf{1}) \geq \mathbf{0}$. In particular, we have a larger $\ell$ compared to constraints from $\mathbf{A}_d$.

## 4.6 The main result

We have shown that for two constants $0 < \epsilon_1 \leq \frac{1}{2}$ and $\epsilon_2 = 1$, under mesh constraints (4.2), the matrices $A_{d^*}$, $A^z$, $A^s$ constructed above satisfy $(A_{d^*} + A^z)\mathbf{1} \geq \mathbf{0}$ and $A_a^+ \leq A^z A_{d^*}^{-1} A^s$.

For any fixed $\epsilon_1 > 0$ and $\epsilon_2 = 1$, $A^z$ also has the same sparsity pattern as $A$. Thus if $\ell$ in (4.2) is replaced by $\sup_{0 < \epsilon_1 \leq \frac{1}{2}, \epsilon_2 = 1} \ell(\epsilon_1, \epsilon_2) = 4$, Theorem 3.7 still applies to conclude that $A^{-1} \geq 0$.

**Theorem 4.1.** *The $Q^2$ spectral element method (4.1) has a monotone matrix $\bar{L}_h$ thus satisfies discrete maximum principle under the following mesh constraints:*

$$h_a h_{a-1} \geq \frac{7}{12} max\{h_b^2, h_{b-1}^2\}, \quad h_b h_{b-1} \geq \frac{7}{12} max\{h_a^2, h_{a-1}^2\},$$

$$min\{h_a, h_{a-1}\} \geq \sqrt{\frac{1}{3}} max\{h_b, h_{b-1}\}, \quad min\{h_b, h_{b-1}\} \geq \sqrt{\frac{1}{3}} max\{h_a, h_{a-1},\}$$

(4.3)

*where $h_a, h_{a-1}$ are mesh sizes for x-axis and $h_b, h_{b-1}$ are mesh sizes for y-variable in four adjacent rectangular cells as shown in Figure 2.*

**Remark 2.** The following global constraint is sufficient to ensure (4.3):

$$\frac{25}{32} \leq \frac{h_m}{h_n} \leq \frac{32}{25},$$

(4.4)

where $h_m$ and $h_n$ are any two grid spacings in a non-uniform grid generated from a non-uniform rectangular mesh for $Q^2$ elements.

**Remark 3.** Though the mesh constraints above may not be sharp, similar constraints are necessary for monotonicity, as will be shown in numerical tests in the next section.

**Remark 4.** For $Q^1$ finite element method solving $-\Delta u = f$ to satisfy discrete maximum principle on non-uniform rectangular meshes [5], the mesh constraints are

$$h_a h_{a-1} \geq \frac{1}{2} max\{h_b^2, h_{b-1}^2\} \quad h_b h_{b-1} \geq \frac{1}{2} max\{h_a^2, h_{a-1}^2\}.$$

(4.5)

# 5 Numerical Tests

## 5.1 Accuracy tests

We show some accuracy tests of the $Q^2$ spectral element method for solving $-\Delta u = f$ on a square $(0,1) \times (0,1)$ with Dirichlet boundary conditions. This scheme is fourth order accurate in $\ell^2$-norm over quadrature points on uniform meshes [15]. On a quasi-uniform mesh, we test the error in $\ell^\infty$-norm to show that this is indeed a high order accurate scheme, which is at least third order accurate. We remark that $Q^2$ spectral element method as a finite difference scheme in $\ell^\infty$ norm is not fourth order accurate even on a uniform mesh, due to the singularity in Green's function in multiple dimensions, see numerical results in [15] and references therein.

Quasi-uniform meshes were generated by setting each pair of consecutive finite element cells along the axis to have a fixed ratio $\frac{h_k}{h_{k-1}} = 1.01$. The scheme is tested for the following very smooth solutions:

1. The Laplace equation $-\Delta u = 0$ with Dirichlet boundary conditions and $u(x,y) = log((x+1)^2 + (y+1)^2) + sin(y)e^x$.

2. Poisson equation $-\Delta u = f$ with homogeneous Dirichlet boundary condition:

$$f(x,y) = 13\pi^2 sin(3\pi y)sin(2\pi x) + 2y(1-y) + 2x(1-x)$$
$$u(x,y) = sin(3\pi y)sin(2\pi x) + xy(1-x)(1-y)$$

$$(5.1)$$

3. Poisson equation $-\Delta u = f$ with nonhomogeneous Dirichlet boundary condition:

$$f = 74\pi^2 cos(5\pi x)cos(7\pi y) - 8$$
$$u = cos(5\pi x)cos(7\pi y) + x^2 + y^2$$

$$(5.2)$$

The errors of $Q^2$ spectral element method on quasi uniform rectangular meshes are listed in Table 1.

## 5.2 Necessity of Mesh Constraints

Even though the mesh constraints derived in the previous section are only sufficient conditions for monotonicity, in practice a mesh constraint is still necessary for the inverse positivity to hold. Consider a non-uniform $Q^2$ mesh with $5 \times 5$ cells on the domain $[0,1] \times [0,1]$, which has a $9 \times 9$ grid for the interior of the domain. Let the mesh on both axes be the same and let the four outer-most cells for each dimension be identical with length $2h$. Then the middle cell has size $2h' \times 2h'$ with $h' = \frac{1}{2} - 2h$. Let the ratio $h'/h$ increase gradually from $h'/h = 1$ (a uniform mesh) until the minimum value of the inverse of the matrix becomes negative. Increasing by values of 0.05, we obtain the first negative entry of $\bar{L}_h^{-1}$ at $h'/h = 5.35$ with $h = 0.0535$ and $h' = 0.2861$, and such a mesh is shown in Figure 4 (a). Figure 4 (b) shows how the smallest entry of $\bar{L}_h^{-1}$ decreases as $h'/h$ increases.

## 6 Concluding remarks

By verifying a relaxed Lorenz's condition, we have discussed suitable mesh constraints, under which the $Q^2$ spectral element method on quasi-uniform meshes is monotone. Even though the derived mesh constraints may not be sharp, a similar constraint is necessary for the monotonicity to hold.

## Acknowledgements

## A Appendix: M-matrices

Nonsingular M-matrices are inverse-positive matrices. There are many equivalent definitions or characterizations of M-matrices, see [20]. The following is a convenient sufficient but not necessary characterization of nonsingular M-matrices [14]:

Table 1: Accuracy test on quasi-uniform meshes.

| Finite Difference Grid | Ratio $\frac{h_i}{h_{i-1}}$ | $Q^2$ spectral element method | |
| --- | --- | --- | --- |
| | | $l^\infty$ error | order |
| test on $-\Delta u = 0$ | | | |
| $7 \times 7$ | 1.01 | 2.66E-5 | - |
| $15 \times 15$ | 1.01 | 1.97E-6 | 3.74 |
| $31 \times 31$ | 1.01 | 1.54E-7 | 3.67 |
| $63 \times 63$ | 1.01 | 1.37E-8 | 3.49 |
| test on (5.1) | | | |
| $7 \times 7$ | 1.01 | 4.92E-2 | - |
| $15 \times 15$ | 1.01 | 3.19E-3 | 3.94 |
| $31 \times 31$ | 1.01 | 2.29E-4 | 3.79 |
| $63 \times 63$ | 1.01 | 1.80E-5 | 3.67 |
| test on (5.2) | | | |
| $7 \times 7$ | 1.01 | 1.20E-0 | - |
| $15 \times 15$ | 1.01 | 1.03E-1 | 3.54 |
| $31 \times 31$ | 1.01 | 9.10E-3 | 3.50 |
| $63 \times 63$ | 1.01 | 9.64E-4 | 3.23 |
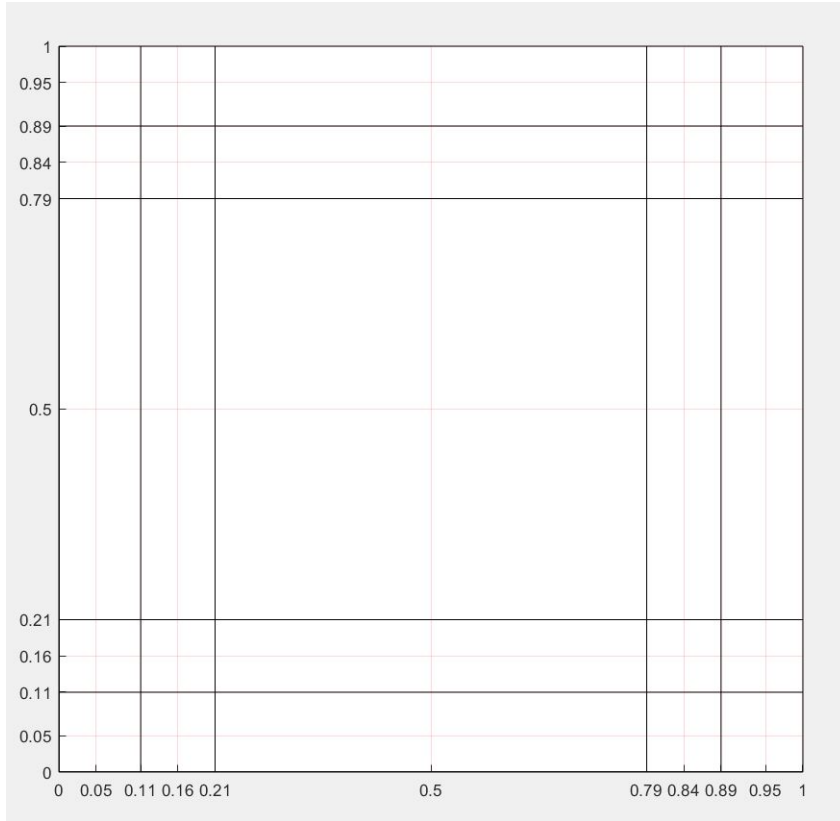
(a) A non-uniform mesh with $5\times5$ cells on which the $Q^2$ spectral element method is no longer monotone. The minimum value of $\bar{L}_h^{-1}$ is $-6.14E-8$.
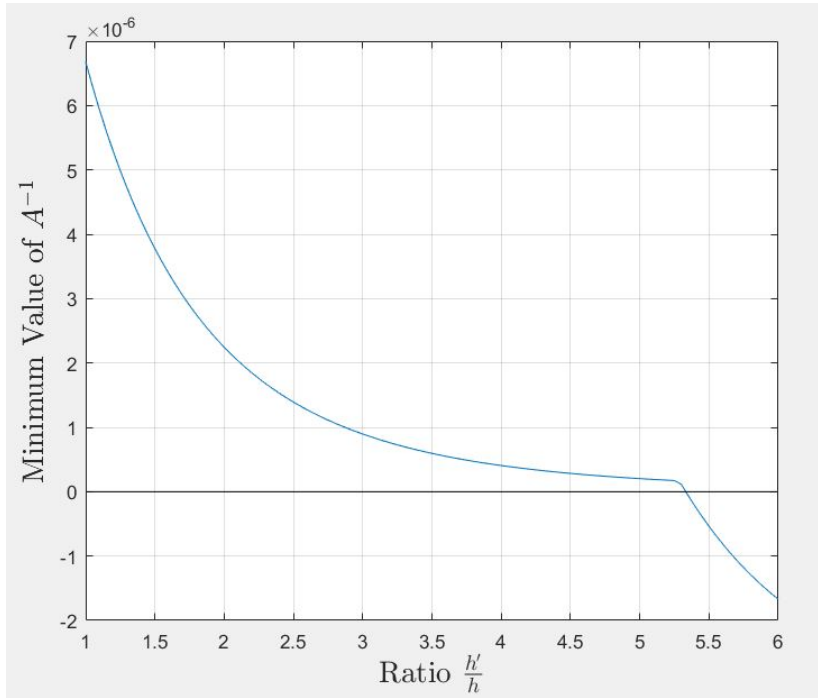


(b) A plot of the minimum value of $\bar{L}_h^{-1}$ as $h'/h$ increases.

Figure 4: Necessity of mesh constraints for inverse positivity $\bar{L}_h^{-1} \geq 0$ where $\bar{L}_h$ is the matrix in $Q^2$ spectral element method on non-uniform meshes.

**Theorem A.1.** *For a real square matrix A with positive diagonal entries and non-positive off-diagonal entries, A is a nonsingular M-matrix if all the row sums of A are non-negative and at least one row sum is positive.*

By condition $K_{35}$ in [20], a sufficient and necessary characterization is,

**Theorem A.2.** *For a real square matrix A with positive diagonal entries and non-positive off-diagonal entries, A is a nonsingular M-matrix if and only if that there exists a positive diagonal matrix D such that AD has all positive row sums.*

**Remark 5.** Non-negative row sum is not a necessary condition for M-matrices. For instance, the following matrix $A$ is an M-matrix by Theorem A.2:

$$A = \begin{bmatrix} 10 & 0 & 0 \\ -10 & 2 & -10 \\ 0 & 0 & 10 \end{bmatrix}, D = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}, AD = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 4 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

# Acknowledgments

**References**

[1] Erich Bohl and Jens Lorenz. Inverse monotonicity and difference schemes of higher order. a summary for two-point boundary value problems. *Aequationes Mathematicae*, 19(1):1–36, 1979.

[2] James H Bramble and Bert E Hubbard. New monotone type approximations for elliptic problems. *Mathematics of Computation*, 18(87):349–367, 1964.

[3] JH Bramble and BE Hubbard. On the formulation of finite difference analogues of the Dirichlet problem for Poisson's equation. *Numerische Mathematik*, 4(1):313–327, 1962.

[4] JH Bramble and BE Hubbard. On a finite difference analogue of an elliptic boundary problem which is neither diagonally dominant nor of non-negative type. *Journal of Mathematics and Physics*, 43(1-4):117–132, 1964.

[5] I Christie and C Hall. The maximum principle for bilinear elements. *International Journal for Numerical Methods in Engineering*, 20(3):549–553, 1984.

[6] Philippe G Ciarlet. Discrete maximum principle for finite-difference operators. *Aequationes Mathematicae*, 4(3):338–352, 1970.

[7] Lothar Collatz. *The numerical treatment of differential equations*. Springer-Verlag, Berlin, 1960.

[8] Logan J. Cross and Xiangxiong Zhang. On the monotonicity of $Q^3$ spectral element method for Laplacian. *arXiv:2010.07282*, 2023.

[9] Werner Höhn and Hans Detlef Mittelmann. Some remarks on the discrete maximum-principle for finite elements of higher order. *Computing*, 27(2):145–154, 1981.

[10] Jingwei Hu and Xiangxiong Zhang. Positivity-preserving and energy-dissipative finite difference schemes for the Fokker-Planck and Keller-Segel equations. *IMA Journal of Numerical Analysis*, 43(3):1450–1484, 2023.

[11] Vladimir Ivanovitch Krylov and Leonid Vital'evitch Kantorovitch. *Approximate methods of higher analysis*. P. Noordhoff, 1958.

[12] Hao Li, Daniel Appelö, and Xiangxiong Zhang. Accuracy of spectral element method for wave, parabolic, and Schrödinger equations. *SIAM Journal on Numerical Analysis*, 60(1):339–363, 2022.

[13] Hao Li, Shusen Xie, and Xiangxiong Zhang. A high order accurate bound-preserving compact finite difference scheme for scalar convection diffusion equations. *SIAM Journal on Numerical Analysis*, 56(6):3308–3345, 2018.

[14] Hao Li and Xiangxiong Zhang. On the monotonicity and discrete maximum principle of the finite difference implementation of $C^0$-$Q^2$ finite element method. *Numerische Mathematik*, 145:437–472, 2020.

[15] Hao Li and Xiangxiong Zhang. Superconvergence of high order finite difference schemes based on variational formulation for elliptic equations. *Journal of Scientific Computing*, 82(2):36, 2020.

[16] Hao Li and Xiangxiong Zhang. A high order accurate bound-preserving compact finite difference scheme for two-dimensional incompressible flow. *Communications on Applied Mathematics and Computation*, pages 1–29, 2023.

[17] Chen Liu, Yuan Gao, and Xiangxiong Zhang. Structure preserving schemes for Fokker-Planck equations of irreversible processes. *Journal of Scientific Computing*, 98(1):4, 2024.

[18] Chen Liu and Xiangxiong Zhang. A positivity-preserving implicit-explicit scheme with high order polynomial basis for compressible Navier–Stokes equations. *Journal of Computational Physics*, 493:112496, 2023.

[19] Jens Lorenz. Zur inversmonotonie diskreter probleme. *Numerische Mathematik*, 27(2):227–238, 1977.

[20] Robert J Plemmons. M-matrix characterizations. I—-nonsingular M-matrices. *Linear Algebra and its Applications*, 18(2):175–188, 1977.

[21] Jie Shen and Xiangxiong Zhang. Discrete maximum principle of a high order finite difference scheme for a generalized Allen-Cahn equation. *Commun. Math. Sci.*, 20(5):1409–1436, 2022.

[22] M Sulman and T Nguyen. A positivity preserving moving mesh finite element method for the Keller–Segel chemotaxis model. *Journal of Scientific Computing*, 80(1):649–666, 2019.

[23] JR Whiteman. Lagrangian finite element and finite difference methods for poisson problems. In *Numerische Behandlung von Differentialgleichungen*, pages 331–355. Springer, 1975.

[24] Jinchao Xu and Ludmil Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Mathematics of Computation*, 68(228):1429–1446, 1999.